Rate-Constrained Shaping Codes for Finite-State Channels With Cost

Yi Liu *Member, IEEE,*[*] **Yonglong Li** *Member, IEEE,*[*‡] **Pengfei Huang** *Member, IEEE,*[*]
Paul H. Siegel *Life Fellow, IEEE*[*]
[*]Department of Electrical and Computer Engineering, CMRR, University of California, San Diego, La Jolla, CA
[‡]National University of Singapore
yiliu0812@gmail.com, elelong@nus.edu.sg, pehuangucsd@gmail.com, psiegel@ucsd.edu

2022 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-2159-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISIT50566.2022.9834441

Abstract—**Shaping codes are used to generate code sequences in which the symbols obey a prescribed probability distribution. They arise naturally in the context of source coding for noiseless channels with unequal symbol costs. Recently, shaping codes have been proposed to extend the lifetime of flash memory and reduce DNA synthesis time. In this paper, we study a general class of shaping codes for noiseless finite-state channels with cost and i.i.d. sources. We establish a relationship between the code rate and minimum average symbol cost. We then determine the rate that minimizes the average cost per source symbol (total cost). An equivalence is established between codes minimizing average symbol cost and codes minimizing total cost, and a separation theorem is proved, showing that optimal shaping can be achieved by a concatenation of optimal compression and optimal shaping for a uniform i.i.d. source.**

I. Introduction

Shaping codes are used to encode information for use on channels with symbol costs under an average cost constraint. They find application in data transmission with a power constraint, where constellation shaping is achieved by addressing into a suitably designed multidimensional constellation or, equivalently, by incorporating, either explicitly or implicitly, some form of non-equiprobable signaling. More recently, shaping codes have been proposed for use in data storage applications: coding for flash memory to reduce device wear [17], and coding for efficient DNA synthesis in DNA-based storage [14]. Motivated by these applications, [18] investigated information-theoretic properties and design of rate-constrained fixed-to-variable length shaping codes for memoryless noiseless channels with cost and general i.i.d. sources. In this paper, we extend the results in [18] to rate-constrained shaping codes for finite-state noiseless channels with cost and general i.i.d. sources.

Finite-state noiseless channels with cost trace their conceptual origins to Shannon's 1948 paper that launched the study of information theory [23]. In that paper, Shannon considered the problem of transmitting information over a telegraph channel. The telegraph channel is a finite-state graph and the channel symbols – dots and dashes – have different time durations, which can be interpreted as integer transmission costs. Shannon defined the combinatorial capacity of this channel and gave an explicit formula. He also determined the symbol probabilities that maximize the entropy per unit cost, and showed the equivalence of this probabilistic definition of capacity to the combinatorial capacity. In [4], this result was then generalized to arbitrary non-negative symbol costs. In [11], a new proof technique for deriving the combinatorial capacity was introduced for non-integer costs and another proof of the equivalence of combinatorial and probabilistic definitions of capacity was given. In [2] and [3], a generating function approach was used to extend the equivalence to a larger class of constrained systems.

We refer to the problem of designing codes that achieve the capacity, i.e., that maximize the information rate per unit cost, or, equivalently, that minimize the cost per information bit, as the *type-II shaping problem*. Several researchers have considered this problem. In [4], modified Shannon-Fano codes, based on matching the probability of source and codeword sequences, were introduced, and they were shown to be asymptotically optimal. A similar idea was used in [22], where an arithmetic coding technique was introduced. Several works extend coding algorithms for memoryless channels to finite-state channels. In [2], a finite-state graph was transformed to its memoryless representation $\mathcal{M}$ and a normalized geometric Huffman code was used to design a asymptotically capacity achieving code on $\mathcal{M}$. In [8], the author extended the dynamic programming algorithm introduced in [7] to finite-state channels. The proposed algorithm finds locally optimal codes for each starting state, but the algorithm does not guarantee global optimality. In [6], an iterative algorithm that can find globally optimal codes was proposed.

The concepts of combinatorial capacity and probabilistic capacity can be generalized to the setting where there is a constraint on the average cost per transmitted channel symbol. The probabilistic capacity was determined in [20] and [9], where the entropy-maximizing stationary Markov chain satisfying the average cost constraint was found. The relationship between cost-constrained combinatorial capacity and probabilistic capacity was also addressed in [10]. The equivalence of the two definitions of cost-constrained capacity was proved in [25], and an alternative proof was recently given in [15], where methods of analytic combinatorics in several variables were used to directly evaluate the cost-constrained combinatorial capacity.

We refer to the problem of designing codes that achieve the cost-constrained capacity, i.e., that minimize average cost per symbol for a given code rate, as the *type-1 shaping problem*. This problem has also been addressed by several authors. In [10], an asymptotically optimal block code was introduced by considering codewords that start and end at the same state. In [12], the authors construct fixed-to-fixed length and variable-to-fixed length codes based on state-splitting methods [1] for magnetic recording and constellation shaping applications. Other constructions can be found in [13], [24] and [26].

In this paper, we address the problem of designing shaping codes for noiseless finite-state channels with cost and general i.i.d. sources. We systematically study the fundamental properties of these codes from the perspective of symbol distribution, average cost, and entropy rate using the theory of finite-state word-valued sources. We derive fundamental bounds relating these quantities and establish an equivalence between optimal type-I and type-II shaping codes. A generalization of Varn coding [28] is shown to provide an asymptotically optimal type-II shaping code for uniform i.i.d. sources. Finally, we prove separation theorems showing that optimal shaping for a general i.i.d. source can be achieved by a concatenation of optimal lossless compression with an optimal shaping code for a uniform i.i.d. source.

In Section II, we define finite-state channels with cost and review the combinatorial and probabilistic capacities associated with the type-I and type-II shaping problems. In Section III, we define finite-state variable length codes for channels with cost and characterize properties of the codeword process using the theory of finite-state word-valued sources. In Section IV, we analyze codes for the type-I shaping problem. We develop a theoretical bound on the trade-off between the rate – or more precisely, the corresponding *expansion factor* – and the average cost of a type-I shaping code. We then study codes for the type-II shaping problem. We derive the relationship between the code expansion factor and the total cost and determine the optimal expansion factor. In Section V, we consider the problem of designing optimal shaping codes. We prove an equivalence theorem showing that both type-I and type-II shaping codes can be realized using a type-II shaping code for a channel with modified edge cost. Using a generalization of Varn coding [28], we propose an asymptotically optimal type-II shaping code on this modified channel for a uniform i.i.d. source. We then extend our construction to arbitrary i.i.d. sources by introducing a separation theorem, which states that optimal shaping can be achieved by a concatenation of lossless compression and optimal shaping for a uniform i.i.d. source.

Due to space constraints, we must omit many detailed proofs, which can be found in [16]. However, we remark that several new proof techniques are required to extend the results on block shaping codes for memoryless channels with cost in [18] to the corresponding results on finite-state shaping codes for finite-state channels with cost in this paper.

II. Noiseless Finite-State Costly Channel

Let $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ be an irreducible finite directed graph, with vertices $\mathcal{V}$ and edges $\mathcal{E}$. Each edge e has an *initial state* $\sigma(e) \in \mathcal{V}$ and a *terminal state* $\tau(e) \in \mathcal{V}$. A *finite-state costly channel* is a noiseless channel with cost associated with $\mathcal{H}$, where each edge $e \in \mathcal{E}$ is assigned a non-negative cost $w(e) \geq 0$. We assume that between any pair of vertices $(v, v') \in \mathcal{V} \times \mathcal{V}$, there is at most one edge. If not, we can always convert it to another graph that satisfies this condition by state splitting [19]. An example of such a channel is given in Example 1.

Example 1. In SLC NAND flash memory, cells are arranged in a grid and programming a cell affects its neighbors. One example of this phenomenon is *inter-cell inference* (ICI) [27]. Cells have two states: *programmed*, corresponding to bit 1, and *erased*, corresponding to bit 0. Due to ICI, programming a cell will damage its neighbors cells. Each length-3 sequence has a cost associated with the damage to the middle bit, as shown in Table I. We can convert this table into a directed graph with vertices $\mathcal{V}=\{00, 01, 10, 11\}$, as shown in Fig. 1, where the the edge from vertex ab to vertex bc corresponds to sequence abc.

TABLE I: Flash memory channel cost.

e	000	001	010	011	100	101	110	111
$w(e)$	1	2	4	4	2	3	4	4

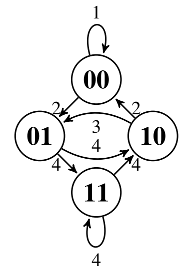

Fig. 1: Flash memory channel

A. Channel capacity with average cost constraint

Given a length-n edge sequence e_1^n, the cost of this sequence is defined as $W(e_1^n) = \sum_{i=1}^n w(e_i)$, and the average cost of this sequence is defined as $A(e_1^n) = \frac{1}{n}W(e_1^n)$. If $K_n(W)$ is the number of sequences of length-n with average cost less than or equal to W, then the *combinatorial capacity* for a given average cost constraint [25], or *cost-constrained capacity*, is

$$C_{I,\text{comb}}(W) \stackrel{\text{def}}{=} \limsup_{n\to\infty} \frac{1}{n} \log_2 |K_n(W)|. \tag{1}$$

We also refer to this definition as *type-I combinatorial capacity*.

Let $\mathbf{E}$ be a stationary Markov process with entropy rate $H(\mathbf{E})$ and average cost $A(\mathbf{E})$. The *type-I probabilistic capacity* for a given average cost constraint W, or *cost-constrained probabilistic capacity*, is

$$C_{I,\text{prob}}(W) \stackrel{\text{def}}{=} \sup_{\mathbf{E}:A(\mathbf{E})\leq W} H(\mathbf{E}). \tag{2}$$

The maxentropic Markov chain for a given W was derived in [9] and [20]. The result relies on the one-step cost-enumerator matrix $D(S)$, where $S \geq 0$, with entries

$$d_{vv'}(S) = \begin{cases} 2^{-Sw(e)}, & \text{if } \exists \text{ s.t. } \sigma(e){=}v,\ \tau(e){=}v', \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Denote by $\lambda(S)$ its Perron root and by vectors $E_L = [P_v/\rho_v]$ and $E_R = [\rho_v]^\top$ the corresponding left and right eigenvectors of $D(S)$ such that $E_L E_R = 1$. Given an average cost constraint $W(S)$ the maxentropic Markov chain has transition probabilities

$$P_e(S) = \frac{1}{\rho_{\sigma(e)}} \frac{2^{-Sw(e)}}{\lambda(S)} \rho_{\tau(e)} \tag{4}$$

such that

$$W(S) = \frac{1}{\lambda(S)} \sum_{e\in\mathcal{E}} w(e) P_{\sigma(e)} 2^{-Sw(e)} \frac{\rho_{\tau(e)}}{\rho_{\sigma(e)}}, \tag{5}$$

and the type-I probabilistic capacity of this channel is

$$C_{I,\text{prob}}(W(S)) = \log_2 \lambda(S) + SW(S). \tag{6}$$

It was shown in [25], [15] that $C_{I,\text{comb}}(W) = C_{I,\text{prob}}(W)$.

B. Channel capacity without cost constraint

Denote by $K(W)$ the number of distinct sequences e^* with cost equal to W. The *combinatorial capacity*, or the *type-II combinatorial capacity*, of this channel is defined as

$$C_{II,\text{comb}} \stackrel{\text{def}}{=} \limsup_{W\to\infty} \frac{1}{W} \log_2 K(W). \tag{7}$$

Similarly, the *type-II probabilistic capacity* of this channel is defined as

$$C_{II,\text{prob}} \stackrel{\text{def}}{=} \sup_{\mathbf{E}} \frac{H(\mathbf{E})}{A(\mathbf{E})}. \tag{8}$$

In [11], it was proved that the transition probabilities of the maxentropic Markov process are $P_e(S_0)$, where S_0 satisfies $\lambda(S_0) = 1$. It was also proved that

$$C_{II,\text{comb}} = C_{II,\text{prob}} = S_0. \tag{9}$$

In [2] and [3], the equivalence between $C_{II,\text{comb}}$ and $C_{II,\text{prob}}$ was extended to a larger class of constrained systems.

III. Finite-State Variable-Length Codes: A Word-Valued Source Approach

A. Finite-State Variable-Length Codes

Let $\mathbf{X} = X_1X_2\ldots$, where $X_i \sim P(X)$ for all i, be an i.i.d. source with a final size alphabet $\mathcal{X}$. Let $|\mathcal{X}|$ denote the size of the alphabet and $P(x^*)$ denote the probability of any finite sequence x^*. A finite-state variable-length code on graph $\mathcal{H}$ is a mapping $\phi : \mathcal{V} \times \mathcal{X}^q \to \mathcal{E}^*$. For simplicity and without loss of generality, we assume $q = 1$ and define its codebook as $\mathcal{Y} = \{y = \phi(v,x) | v \in \mathcal{V}, x \in \mathcal{X}\}$. We also use $\sigma(y)$ and $\tau(y)$ to denote the initial and terminal states of $y \in \mathcal{Y}$, and denote its length by $L(y)$. We assume this mapping has the following two properties:

- The subcodebook $\mathcal{Y}_v = \{\phi(v,x) | x \in \mathcal{X}\}$ is prefix-free for all $v \in \mathcal{V}$.
- $L_{\min} \le L(y) \le L_{\max}$ for all $y \in \mathcal{Y}$, for some $0 < L_{\min} \le L_{\max}$.

Starting from vertex v_0, and given an input sequence $x_1, x_2, \ldots,$ such that $x_i \in \mathcal{X}$, ϕ generates a sequence of of variable-length codewords $y_1, y_2, \ldots,$ such that $y_i \in \mathcal{Y}$, that satisfies the following constraints.

$$\begin{aligned} y_i &= \phi(v_{i-1}, x_i), \\ v_i &= \tau(y_i) = \sigma(y_{i+1}). \end{aligned} \tag{10}$$

Define mappings $F : \mathcal{Y} \times \mathcal{X} \to \mathcal{Y}$, and $G : \mathcal{V} \times \mathcal{X} \to \mathcal{V}$ as $F(y,x) = \phi(\tau(y), x)$, $G(v,x) = \tau(\phi(v,x))$, respectively. From (10), we have $y_i = F(y_{i-1}, x_i)$ and $v_i = G(v_{i-1}, x_i)$, suggesting the definition of *codeword graph* $\mathcal{F}_0$ and *state graph* $\mathcal{G}_0$, as follows.

- Codeword graph $\mathcal{F}_0 = (\mathcal{V}_{\mathcal{F}_0}, \mathcal{E}_{\mathcal{F}_0})$:
  - Vertices $\mathcal{V}_{\mathcal{F}_0} = \mathcal{Y}$
  - Edges $\mathcal{E}_{\mathcal{F}_0} = \{e = (y, y') | y' = F(y,x), x \in \mathcal{X}\}$.
- State graph $\mathcal{G}_0 = (\mathcal{V}_{\mathcal{G}_0}, \mathcal{E}_{\mathcal{G}_0})$:
  - Vertices $\mathcal{V}_{\mathcal{G}_0} = \mathcal{V}$.
  - Edges $\mathcal{E}_{\mathcal{G}_0} = \{e = (v, v') | v' = G(v,x),\ x \in \mathcal{X}\}$.

We now choose irreducible subgraphs of $\mathcal{F}_0$ and $\mathcal{G}_0$. Let $\mathcal{G} \subseteq \mathcal{G}_0$ be the irreducible component of $\mathcal{G}_0$ that contains v_0. Its state and edge sets are denoted by $\mathcal{V}_\mathcal{G}$ and $\mathcal{E}_\mathcal{G}$, respectively. It can be shown that, given an i.i.d. source $\mathbf{X}$ and mapping G, we obtain a Markov chain with transition probabilities

$$t^{\mathcal{G}}_{vv'} = \sum_{x:G(v,x)=v'} P(x). \tag{11}$$

We denote by $\pi_\mathcal{G}$ the stationary distribution of this Markov chain. We define $\mathcal{F} \subset \mathcal{F}_0$ as follows. Its vertex set $\mathcal{V}_\mathcal{F}$ is

$$\mathcal{V}_\mathcal{F} = \{y = \phi(v,x) \in \mathcal{Y} | v \in \mathcal{V}_\mathcal{G}, x \in \mathcal{X}\} \tag{12}$$

and its edge set $\mathcal{E}_\mathcal{F}$ is

$$\mathcal{E}_\mathcal{F} = \{(y, y') | y' = F(y,x), x \in \mathcal{X}\} \subseteq \mathcal{V}_\mathcal{F} \times \mathcal{V}_\mathcal{F}. \tag{13}$$

We have the following lemma.

Lemma 1. *The graph $\mathcal{F}$ is irreducible. Given a distribution $P(x)$ and mapping F, we obtain a Markov chain with transition probabilities*

$$P(y'|y) = P(x),\ \text{s.t. } F(y,x) = y'. \tag{14}$$

The stationary distribution of this Markov chain is

$$\pi_\mathcal{F}(y) = \pi_\mathcal{G}(v) P(x),\ \text{s.t. } \phi(v,x) = y. \tag{15}$$

□

Using the law of large numbers for irreducible Markov chains [5, Exercise 5.5], [19, Theorem 3.21] and the dominated convergence theorem, we know that the expected length of the codeword process $\mathbf{Y}$ is

$$\mathbb{E}(L) \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\left(\sum_{m=1}^{n} L(Y_m) \right) = \sum_{y \in \mathcal{V}_\mathcal{F}} L(y) \pi_\mathcal{F}(y). \tag{16}$$

Given an edge $e \in \mathcal{E}$ and codeword $y = \phi(v,x)$, we denote by $N_e(y)$ the total number of occurrences of e in y. We can similarly prove that

$$\mathbb{E}(N_e) = \sum_{y \in \mathcal{V}_\mathcal{F}} N_e(y) \pi_\mathcal{F}(y). \tag{17}$$

B. Finite-State Word-Valued Source

Introduced in [21], a *word-valued source* is a discrete random process that is formed by sequentially encoding the symbols of an i.i.d. random process $\mathbf{X}$ into corresponding codewords over an alphabet $\mathcal{E}$. In this paper, the mapping is a function of both input symbols and the starting state. We refer to the process $\mathbf{E}$, formed by an i.i.d. source process and mapping ϕ, as a *finite-state word-valued source*.

Given an encoded sequence $e_1e_2\ldots$, the probability of sequence e_1^n is $Q(e_1^n)$, and the number of occurrences of e is $N_e(e_1^n)$. The following properties of the process $\mathbf{E}$ are of interest.

- The asymptotic symbol occurrence probability
$$\hat{P}_e = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}(N_e(E_1^n)). \tag{18}$$
- The asymptotic average cost
$$A(\phi) = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}(W(E_1^n)). \tag{19}$$
- The entropy rate
$$H(\mathbf{E}) = \lim_{n\to\infty} \frac{1}{n} H(E_1^n). \tag{20}$$

We can prove the following lemmas.

Lemma 2. *For a finite-state code $\phi : \mathcal{V} \times \mathcal{X}^q \to \mathcal{E}^*$ associated with graph $\mathcal{H}$ and input distribution $P(x)$ such that $\mathbb{E}(N_e) < \infty$ for all $e \in \mathcal{E}$ and $\mathbb{E}(L) < \infty$, the asymptotic probability of occurrence $\hat{P}_e$ of this mapping is*

$$\hat{P}_e = \mathbb{E}(N_e) \frac{1}{\mathbb{E}(L)}. \tag{21}$$

The asymptotic average cost $A(\phi)$ of this mapping is

$$A(\phi) = \sum_e \hat{P}_e w(e). \tag{22}$$

□

Lemma 3. *For a finite-state code $\phi : \mathcal{V} \times \mathcal{X}^q \to \mathcal{E}^*$ associated with graph $\mathcal{H}$ and input distribution $P(X)$ such that $H(\mathbf{X}) < \infty$ and $\mathbb{E}(L) < \infty$, the entropy rate of this finite-state word-valued source is*

$$H(\mathbf{E}) = \frac{qH(\mathbf{X})}{\mathbb{E}(L)} = \frac{H(\mathbf{X})}{f}. \tag{23}$$

Here $f = \mathbb{E}(L)/q$ is the expansion factor of the mapping ϕ. □

C. Asymptotic normalized KL-divergence

Similar to the definition of $\hat{P}_e$, the asymptotic probability of occurrence of state $v \in \mathcal{V}$ is defined as

$$\hat{P}_v = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}(N_v(E_1^n)), \tag{24}$$

and we can prove that

$$\hat{P}_v = \sum_{e:\sigma(e)=v} \hat{P}_e \tag{25}$$

Consider a finite-order Markov process $\hat{\mathbf{E}}$ associated with graph $\mathcal{H}$ and transition probabilities

$$t_{vv'}^{\mathcal{H}} = \hat{P}_e/\hat{P}_v. \quad v = \sigma(e), v' = \tau(e) \tag{26}$$

Denote by $\hat{P}(e_1^n)$ the probability of a length-n sequence generated by this process. To measure the difference between $\hat{\mathbf{E}}$ and $\mathbf{E}$, we define the asymptotic normalized KL-divergence

$$\lim_{n\to\infty} \frac{1}{n} D(E_1^n||\hat{E}_1^n) = \lim_{n\to\infty} \sum_{e_1^n\in\mathcal{E}^n} Q(e_1^n)\log_2 \frac{Q(e_1^n)}{\hat{P}(e_1^n)}. \tag{27}$$

The relationship between processes $\mathbf{E}$ and $\hat{\mathbf{E}}$ is summarized in the following lemma.

Lemma 4. *The asymptotic normalized KL-divergence between processes $\mathbf{E}$ and $\hat{\mathbf{E}}$ satisfies*

$$\lim_{n\to\infty} \frac{1}{n} D(E_1^n||\hat{E}_1^n) = H(\hat{\mathbf{E}}) - H(\mathbf{E}) = H(\hat{\mathbf{E}}) - \frac{H(\mathbf{X})}{f}. \tag{28}$$

□

Remark 1. When $H(\hat{\mathbf{E}}) = H(\mathbf{E})$, $\lim_{n\to\infty} \frac{1}{n} D(E_1^n||\hat{E}_1^n) = 0$. Therefore, the codeword process $\mathbf{E}$ approximates the stationary Markov process $\hat{\mathbf{E}}$, in the sense that the asymptotic normalized KL-divergence between $\mathbf{E}$ and $\hat{\mathbf{E}}$ converges to 0.

IV. Optimal Shaping Code Characterization

In this section, we first analyze finite-state codes that minimize the average cost with a given expansion factor. We refer to these as *optimal type-I shaping codes*. We solve the following optimization problem for the optimal asympototic probabilities.

$$\begin{aligned} \underset{\hat{P}_e}{\text{minimize}} \quad & \sum_e \hat{P}_e w(e) \\ \text{subject to} \quad & H(\hat{\mathbf{E}}) = -\sum_e \hat{P}_e \log_2 \frac{\hat{P}_e}{\hat{P}_{\sigma(e)}} \geq \frac{H(\mathbf{X})}{f} \\ & \sum_{\sigma(e)=v} \hat{P}_e = \sum_{\tau(e')=v} \hat{P}_{e'} \text{ and } \sum_e \hat{P}_e = 1. \end{aligned} \tag{29}$$

In [15], the authors discuss *cost-diverse* and *cost-uniform* graphs. A graph is cost-diverse if it has at least one pair of equal-length paths with different costs that connect the same pair of vertices. Otherwise, it is called cost-uniform. It can be proved that the edge costs $w(e)$ of a cost-uniform graph can be expressed as $w(e) = -\mu(\sigma(e)) + \mu(\tau(e)) - \alpha$, where α is a constant. The following theorem, whose proof uses the KKT conditions, relates to the minimum achievable average cost of a finite-state code.

Theorem 5. *On a cost-diverse graph, the average cost of a finite-state code $\phi : \mathcal{V} \times \mathcal{X}^q \to \mathcal{E}^*$ with expansion factor f is lower bounded by*

$$A_{min}(f) = \sum_e \hat{P}_e w(e) = \frac{H(\mathbf{X})}{Sf} - \frac{\log_2 \lambda(S)}{S} \tag{30}$$

where $\hat{P}_e = \frac{\hat{P}_{\sigma(e)}}{\rho_{\sigma(e)}} \frac{2^{-Sw(e)}}{\lambda(S)} \rho_{\tau(e)}$, $\lambda(S)$ is the Perron root of the matrix $D(S)$, $E_L = [\hat{P}_v/\rho_v]$ and $E_R = [\rho_v]^\top$ are the corresponding eigenvectors such that $E_L E_R = 1$, and S is the constant such that

$$H(\hat{\mathbf{E}}) = -\sum_e \hat{P}_e \log_2 \frac{\hat{P}_e^I}{\hat{P}_{\sigma(e)}} = H(\mathbf{E}) = \frac{H(\mathbf{X})}{f}. \tag{31}$$

On a cost-uniform graph, the average cost for any shaping code is a constant $-\alpha$. □

Remark 2. When the minimum average cost is achieved, we have $H(\hat{\mathbf{E}}) = H(\mathbf{E})$. As shown in Remark 1, the codeword sequence approximates a finite-order stationary Markov process with transition probabilities $\{\hat{P}_e/\hat{P}_{\sigma(e)}\}$. □

The *total cost* of a finite-state code is

$$T(\phi) = \lim_{n\to\infty} \frac{\mathbb{E}(W(\phi(X^{nq})))}{nq} = f\sum \hat{P}_e w(e). \tag{32}$$

We refer to codes that minimize the total cost as *optimal type-II shaping codes*. The optimization problem for the optimal asymptotic probabilities and expansion factor is as follows.

$$\begin{aligned} \underset{\hat{P}_e, f}{\text{minimize}} \quad & f\sum_e \hat{P}_e w(e) \\ \text{subject to} \quad & H(\hat{\mathbf{E}}) \geq H(\mathbf{E}) = \frac{H(\mathbf{X})}{f} \\ & \sum_{\sigma(e)=v} \hat{P}_e = \sum_{\tau(e')=v} \hat{P}_{e'} \text{ and } \sum_e \hat{P}_e = 1. \end{aligned} \tag{33}$$

We have the following theorem that determines the minimum achievable total cost of a finite-state code.

Theorem 6. *If a cost-0 cycle does not exist, the minimum total cost of a finite-state code $\phi : \mathcal{V} \times \mathcal{X}^q \to \mathcal{E}^*$ is given by*

$$T_{min} = f\sum_e \hat{P}_e w(e) = H(\mathbf{X})/S \tag{34}$$

where $\hat{P}_e = \frac{\hat{P}_{v_{\sigma(e)}}}{\rho_{\sigma(e)}} 2^{-Sw(e)} \rho_{\tau(e)}$, S is a constant such that $\lambda(S) = 1$, and $E_L = [\hat{P}_v/\rho_v]$ and $E_R = [\rho_v]^\top$ are the corresponding eigenvectors such that $E_L E_R = 1$. The optimal expansion factor f is

$$f = \frac{H(\mathbf{X})}{-\sum \hat{P}_e \log_2 \frac{\hat{P}_e}{\hat{P}_{v_{\sigma(e)}}}} = \frac{H(\mathbf{X})}{S\sum \hat{P}_e w(e)}. \tag{35}$$

If there is a cost-0 cycle in $\mathcal{H}$, the total cost is a decreasing function of f. □

V. Optimal Shaping Code Design

In this section, we consider the problem of designing optimal type-I and type-II shaping codes.

A. Equivalence Theorems

The next two theorems establish equivalences between type-I and type-II shaping codes.

Theorem 7. *Given a noiseless finite-state costly channel with edge costs $\{w(e)\}$, consider a modified channel with edge costs*

$$\begin{aligned} w'(e) &= -\log_2 \frac{\hat{P}_e^\circ}{\hat{P}_{\sigma(e)}^\circ} \\ &= S^\circ w(e) + \log_2 \rho_{\sigma(e)}^\circ - \log_2 \rho_{\tau(e)}^\circ + \log_2 \lambda(S^\circ) \end{aligned} \tag{36}$$

where $\hat{P}_e^\circ$, $\hat{P}_{\sigma(e)}^\circ$, ρ_e° and S° are calculated in Theorem 5.

For any $\gamma, \eta > 0$, there exists a $\delta > 0$ such that if finite-state code $\phi : \mathcal{V} \times \mathcal{X}^q \to \mathcal{E}^*$ with expansion factor f' and asymptotic probabilities $\hat{P}_e'$ satisfies

$$f'\sum \hat{P}_e' w'(e) - H(\mathbf{X}) < \delta, \tag{37}$$

then the average cost of this code is upper bounded by

$$\sum \hat{P}'_e w(e) - \left(\frac{H(\mathbf{X})}{S^\circ f} - \frac{\log_2 \lambda(S^\circ)}{S^\circ}\right) < \gamma \tag{38}$$

and the expansion factor of this code f' satisfies $|f' - f| < \eta$.

Theorem 8. *Given a noiseless finite-state costly channel that does not contain a cost-0 cycle. Let $S^\star$ and $f^\star$ be the constant and expansion factor calculated in Theorem 6. For any $\gamma > 0$, there exist $\delta, \eta > 0$ such that if a finite-state code $\phi : \mathcal{V} \times \mathcal{X}^q \to \mathcal{E}^*$ with expansion factor f' and asymptotic probabilities $\hat{P}'_e$ satisfies*

$$\sum \hat{P}'_e w(e) - A_{min}(f') < \delta, \quad |f' - f^\star| < \eta \tag{39}$$

then the total cost of this code satisfies

$$f' \sum \hat{P}'_e w(e) - \frac{H(\mathbf{X})}{S^\star} < \gamma. \tag{40}$$

B. Generalized Varn Code

We now describe an asymptotically optimal type-II shaping code for uniform i.i.d. sources based on a generalization of Varn coding [28]. We consider the channel with modified edge costs

$$\begin{aligned} w'(e) &= -\log_2 \hat{P}^\star_e / \hat{P}^\star_{\sigma(e)} \\ &= S^\star w(e) + \log_2 \rho^\star_{\sigma(e)} - \log_2 \rho^\star_{\tau(e)} \end{aligned} \tag{41}$$

where $\hat{P}^\star_e$, $\hat{P}^\star_{\sigma(e)}$, $S^\star$, and $\rho^\star_e$ are given in Theorem 6. It is easy to show that the symbol occurrence probabilities of optimal type-II shaping codes on the channel with costs $\{w'(e)\}$ are identical to those on the original channel with costs $\{w(e)\}$.

Given a uniform i.i.d. input source $\mathcal{X}$, a *generalized Varn code* on the noiseless finite-state costly channel is a collection of tree-based variable-length mappings, $\phi : \mathcal{V} \times \mathcal{X}^q \to \mathcal{E}^*$. The codewords in $\mathcal{Y}_v$ are generated according to the following steps.

1) Set state $v \in \mathcal{V}$ as the root of the tree.
2) Expand the root node. The cost of a leaf node is the cost of the edge from the root node to the leaf node, where the edge costs are the modified costs $\{w'(e)\}$ defined in (41).
3) Expand the leaf node that has the lowest cost. The cost of a leaf node in the resulting tree is the cost of the path from the root node to the leaf node, where the edge costs are $\{w'(e)\}$ defined in (41).
4) Repeat step 3 until the total number of leaf nodes $M \geq |\mathcal{X}|^q$. Delete the leaf nodes that have the largest cost until the number of leaf nodes equals $|\mathcal{X}|^q$. Each path from the root node v to a leaf node represents a codeword in $\mathcal{Y}_v$.

The following lemma gives an upper bound on the total cost of a generalized Varn code.

Lemma 9. *The total cost of a generalized Varn code $\phi : \mathcal{V} \times \mathcal{X}^q \to \mathcal{E}^*$ is upper bounded by*

$$T(\phi) \leq \frac{\log_2 M}{q} + \frac{\max_e\{w'(e)\}}{q} \underset{q\to\infty}{=} \log_2 |\mathcal{X}|. \tag{42}$$

□

Remark 3. By extending some leaf nodes to states that are not visited by the original code, we can make graph $\mathcal{G}_0$ a complete graph. Then we can choose any state as the starting state. This operation only adds a constant to the cost of a codeword and therefore does not affect the asymptotic performance of the generalized Varn code.

When combined with Lemma 9, the following lemma proves that a generalized Varn code is an asympotically optimal type-II shaping code on the channel with costs $\{w(e)\}$.

Lemma 10. *Given a noiseless finite-state costly channel $\{w(e_{ij})\}$. Let $\{w'(e)\}$ be the modified costs defined in (41). If there is a finite-state code $\phi : \mathcal{V} \times \mathcal{X}^q \to \mathcal{E}^*$ such that*

$$f \sum \hat{P}_e w'(e) - H(\mathbf{X}) < \delta \tag{43}$$

for some $\delta > 0$, then the total cost of this code satisfies

$$f \sum \hat{P}_e w(e) - \frac{H(\mathbf{X})}{S^\star} < \frac{\delta}{S^\star}. \tag{44}$$

□

Example 2. For the channel introduced in Example 1, the optimal symbol distributions that minimize the total cost are shown in Table II. Based on the distribution, we can design a generalized Varn code on the channel with modified edge costs shown in Table III. The total cost as a function of codebook size is shown in Fig. 2.

TABLE II: Probabilities for SLC flash channel that minimize total cost.

e	000	001	010	011	100	101	110	111
$\hat{P}_e$	0.4318	0.1323	0.1135	0.0593	0.1323	0.0405	0.0593	0.0310

TABLE III: Modified cost for flash memory channel.

e	000	001	010	011	100	101	110	111
$w'(e)$	0.3805	2.0923	0.6068	1.5423	0.3855	2.0923	0.6068	1.5423

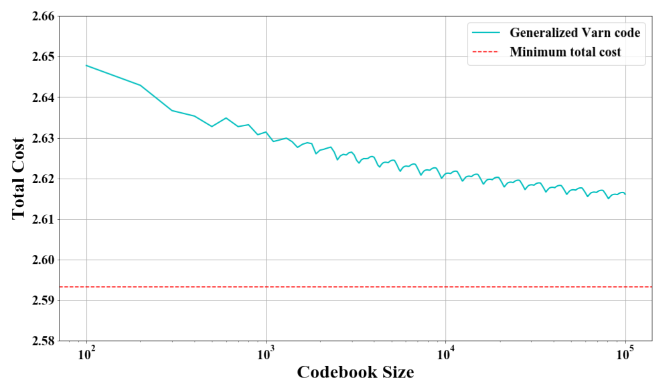

Fig. 2: The total cost of a generalized Varn code on the SLC flash channel.

C. Separation Theorem

We now present a separation theorem for shaping codes. It states that the minimum total cost can be achieved by a concatenation of optimal lossless compression with an optimal shaping code for a uniform i.i.d. source.

Theorem 11. *Given an i.i.d. source* $\mathbf{X}$ *and a noiseless finite-state costly channel with edge costs $\{w(e)\}$, the minimum total cost can be achieved by a concatenation of an optimal lossless compression code with a binary optimal type-II shaping code for a uniform i.i.d. source.*

Theorem 12. *Given the i.i.d. source* $\mathbf{X}$*, the noiseless finite-state costly channel with edge costs $\{w(e)\}$, and the expansion factor f, the minimum average cost can be achieved by a concatenation of an optimal lossless compression code with a binary optimal type-I shaping code for uniform i.i.d. source and expansion factor $f' = \frac{f}{H(\mathbf{X})}$.*

By Theorem 8, the optimal type-I shaping code for uniform i.i.d. source in Theorem 12 can be replaced by a suitable optimal type-II shaping code for uniform i.i.d. source.

References

[1] R. Adler, D. Coppersmith, and M. Hassner, "Algorithms for sliding block codes," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 1, pp. 5–22, Jan. 1983.

[2] G. Böcherer, "Capacity-achieving probabilistic shaping for noisy and noiseless channels", Ph.D. dissertation, RWTH Aachen University, 2012.

[3] G. Böcherer, R. Mathar, V. C. da Rocha Jr., and C. Pimentel,"On the capacity of constrained systems," in *Proc. Int. ITG Conf. Source Channel Coding (SCC)*, 2010.

[4] I. Csiszár, "Simple proofs of some theorems on noiseless channels", *Inf. Contr.*, vol. 14, pp. 285–298, 1969.

[5] R. Durrett, *Probability: Theory and Examples*, 3rd ed. Belmont, CA: Duxbury, 2004.

[6] R. Fujita, K. Iwata, and H. Yamamoto, "An iterative algorithm to optimize the average performance of Markov chains with finite states," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, 2019, pp.1902 - 1906.

[7] M. J. Golin and G. Rote, "A dynamic programming algorithm for constructing optimal prefix-free codes with unequal letter costs," *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 1770–1781, Sep. 1998.

[8] K. Iwata and T. Koyama, "A prefix-free coding for finite-state noiseless channels with small coding delay," in *Proc. 2010 Int. Symp. Inf. Theory & its Appl.*, Taichung, Taiwan, Oct. 2010, pp. 473–477.

[9] J. Justesen and T. Høholdt, "Maxentropic Markov chains," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 4, pp. 665–667, Jul. 1984.

[10] R. Karabed, D. L. Neuhoff, A. Khayrallah, *The capacity of costly noiseless channels*, Research report, IBM Research Division, 1988.

[11] A. Khandekar, R. J. McEliece, and E. Rodemich, "The discrete noiseless channel revisited," in *Proc. 1999 Int. Symp. Commun. Theory and Appl.*, pp. 115-137, 1999.

[12] A. S. Khayrallah and D. L. Neuhoff, "Coding for channels with cost constraints," *IEEE Trans. Inf. Theory*, vol. 42, pp. 854-867, May 1996.

[13] V. Y. Krachkovsky, R. Karabed, S. Yang and B. A. Wilson, "On modulation coding for channels with cost constraints", in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, Ju.-Jul. 2014, pp. 421–425.

[14] A. Lenz, Y. Liu, C. Rashtchian, P. H. Siegel, A. Wachter-Zeh, and E. Yaakobi, "Coding for efficient DNA synthesis," in *Proc. IEEE Int. Symp. Inf. Theory*, Los Angeles, CA, Jun. 2020, pp. 2885-2890.

[15] A. Lenz, S. Melczer, C. Rashtchian, and P. H. Siegel, "Multivariate analytic combinatorics for cost constrained channels and subsequence enumeration", arXiv:2111.06105 [cs.IT], Nov. 2021.

[16] Y. Liu, Y. Li, P. Huang, and P. H. Siegel, "Rate-constrained shaping codes for finite-state channels with cost", arXiv:2205.03731 [cs], May, 2022.

[17] Y. Liu and P. H. Siegel, "Shaping codes for structured data," in *Proc. IEEE Globecom*, Washington, D.C., Dec. 4-8, 2016, pp. 1–5.

[18] Y. Liu, P. Huang, A. W. Bergman, P. H. Siegel, "Rate-constrained shaping codes for structured sources", *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 5261–5281, Aug. 2020.

[19] B. H. Marcus, R.M. Roth, and P.H. Siegel, *An introduction to coding for constrained systems*, Lecture Notes, 2001, available online at: ronny.cswp.cs.technion.ac.il/wp-content/uploads/sites/54/2016/05/chapters1-9.pdf

[20] R. J. McEliece and E. R. Rodemich, "A maximum entropy Markov chain", in *Proc. 17th Conf. Inf. Sci. and Syst.*, Johns Hopkins University, Mar. 1983, pp. 245-248.

[21] M. Nishiara and H. Morita, "On the AEP of word-valued sources," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 1116–1120, May 2000.

[22] S. A. Savari and R. G. Gallager, "Arithmetic coding for finite-state noiseless channels," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 100–107, Jan. 1994.

[23] C. E. Shannon, "A mathematical theory of communication, Part I, Part II," *Bell Syst. Tech. J,* vol. 27, pp. 379–423, 1948.

[24] J. B. Soriaga and P. H. Siegel, "On distribution shaping codes for partial-response channels," in *Proc. 41st Annu. Allerton Conf. on Commun., Control, and Comput.*, (Monticello, IL, USA), pp. 468-477, October 2003.

[25] J. B. Soriaga and P. H. Siegel "On the design of finite-state shaping encoders for partial-response channels," in *Proc. of the 2006 Inf. Theory and Appl. Workshop* (ITA2006), San Diego, CA, USA, Feb. 2006.

[26] J. B. Soriaga and P. H. Siegel, "Near-capacity coding systems for partial-response channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Chicago, IL, USA, p. 267, June 2004.

[27] V. Taranalli, H. Uchikawa and P. H. Siegel, "Error analysis and inter-cell interference mitigation in multi-level cell flash memories," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, Jun. 2015 , pp. 271-276.

[28] B. Varn, "Optimal variable length codes (arbitrary symbol cost and equal code word probability)," *Inform. Contr.*, vol. 19, pp. 289–301, 1971.